

```

1 CCTCGGCGGC CCGCATCTGC CCCCGCGCGC CCGCCCTGAG CCCGCCCCGA
51 CTGGGCAGGC GGGGGAGCCC CTACTTCTCT CCCCCCGGGC GGGGGAGCCG
101 GGGGGCAGCG CCGGAGCCCG GGGGGAGCTC AGCCCCCGCG ACCGGCCGGC
151 CAGGGCAGGG GGCAGCTAGG ACGGCCCCGG TCCAGGTGGA GGCCGCAGAG
201 GGCCCAAGGC AAGCAGAGGC AGCAATGGTT GGTCTTGACG GTGGCTGAGC
251 CCCCAGCCCC TGGAATATGC AGCCCGGGGG AGCCCCAGAC AGCGGCAAGG
301 ACGAGGTGGC GGAGTGGGGC GGGAGGCATG GTCTCCACCT ACCGGGTGGC
351 CGTGCTGGGG GCGCGAGGTG TGGGCAAGAG TGCCATCGTG CGCCAGTTCT
401 TGTACAACGA GTTCAGCGAG GTCTGCGTCC CCACCACCGC CCGCCGCCCT
451 TACCTGCCCTG CTGTCGTCAT GAACGGCCAC GTGCACGACC TCCAGATCCT
501 CGACTTTTCCA CCGATCAGCG CTTTCCCTGT CAATACGCTC CAGGAGTGGG
551 CAGACACCTG CTGCAGGGGA CTCCGGAGTG TCCACGCTA CATCCTGGTC
601 TACGACATCT GCTGCTTTGA CAGCTTTGAG TACGTCAAGA CCATCCGCCA
651 GCAGATCCTG GAGACGAGGG TGATCGGAAC CTCAGAGACG CCCATCATCA
701 TCGTGGGCAA CAAGCGGGAC CTGCAGCGCG GACGCGTGAT CCCGCGCTGG
751 AACGTGTGCG ACCTGGTACG CAAGACCTGG AAGTGGCGCT ACGTGGAAATG
801 CTCGGCCAAG TACAACGGC ACATCCTGCT GCTCTTCAGC GAGCTGCTCA
851 AGAGCGTCCG CTGCGCCCGT TGCAAGCAGG TGCACGCTGC CCTGCGCTTC
901 CAGGGCGCGC TGCGCCGCAA CCGCTGCGCC ATCATGTGAC GCCTGCGCGC
951 CCCTCGGGCT GCACCGGCAC TGGCCGAGCG GAGGGCGGGG CCGTACTGCG
1001 GGGCTGGGGC GGGGAGCGGG CGGGAAATGG AACTGTGACG GTCCCCGCCCT
1051 GAGGCCCTTG CAGCCACGCA CCTCCCGGTG AGAAGCAGAG CGCGAGAGGG
1101 AGCCCTCCGT AACTGCCAG CCTTGCCCTT TGCCCCCGT GCTTCTTGGG
1151 ACAGCCGCCCT TCAGTGTGTG ATTTAGTGCA GTGCCCGGCC CGACCCGCGG
1201 GGGTGCCACA GCCTTTTGGG ATGGGGGTGA GCGTGCAATG GAGGCTGGGG
1251 GTGGCGAGGT GCCGCCTTGG CCGGGCCCCC ACGTGTCTTC TCCAGAATGT
1301 GTCTGTCTTT GCCTGGTGTG TTCCTTTCCC GTGTCCGCCC ACCCCAGCGT
1351 CTGTTGGTAC TTACCTGTCT CACCTACCCCT CCAGTCCCTT CCCAGCTCCG
1401 CTCACAGGGC TCTCATTTTCG TCCATCCCTT TGTCGAGAT CCTGGCAGCT
1451 TCTTTGTGAG GCCAGCCCTT CTGACTGTCA GCACCACCGG CACAGGGCAG
1501 AGATGCGGGT GGGCCAAGGA CCACGATCAA GGGGTCCGGG GGACCGAGGT
1551 CCCAGATCAG TGAGGGGAGA AGGTTGAGCT CTCCGGCTTC CAGGGAGACC
1601 TCCCCGCCCA GCAGCCCCCA GAGACACAAC AACCTACCTT CCAGCCTTAA
1651 CTCGATGGTC CGTCCCTGCC AGGTGCCCCT CACTCTTCCT GACCCCAAAG
1701 CCAGATCACC CCCTGGGTTA AAACTTTTTT TCTTTTTTTT TTTTGGACAG
1751 AGTGTGGAAA GGGAGCCCCC CAAAGGATAG CTTCTTTTTC ATGATGCCAG
1801 GCTCCAGTCC TTTATTCCCT TCTGCATACT GCAATCTGAT CTGTGAGACT
1851 GGGGAATGTT GGGTTCTGGG GTCTGGTCTG GGGCAGGATG GTGCCCAGAA
1901 GGGGGTTAGG TTGTCCCAGT GAAAATTCTG TTGCCCCGTC TCAACCCCAT
1951 CTGACTACCC CAGACTCTGC CTGCCTCAGA TCTCAGACTA TCCTGATTAA
2001 TCTGGGGAAG AACAGAGCCA GGGAAAGAAT GGTGGGGACC CCTGTACTTG
2051 GGGGAGACAC ACCTGCATCT TCCTCCTGCC ACAGATGGAG GCCCTCAGGA
2101 TCTGACACCC TCTTGTCCCA ACACAGTCA GCCCTATACC CTAACCTACT
2151 CCACCCCATT TTCTCCGGCT GCCTGGCCGG GTTTCTACCT CTCGTCACCG
2201 GAGCTGATCA CTGTCACTTT TGTACCGATT TAGAAATAAC AATAATAATG
2251 AAGATTCTAG GAATGGCATG AGGGATTGAT GGGGGACTTG GAGGGAGGGA
2301 CAAGTGGTGC CCTGTCCCTT GCTCCCTTGG CCAAAGAAAG CTGTCTTGA
2351 GGCTGAGCCC TCAGCCCTGG CCTGGTGGGG GGACAGCAAG GTCCCTTGTT
2401 ATAAGAGGGG CAGAGAGGAC AACTCCGCTT TGGCCAACCT AGCCAAGGCT
2451 GCAGCATATA GACCAGGAAA TCAGGTAGCC CAGACTGGTG ATGGAGCAGA
2501 GTCTGGGGGA AGGGTCGTGG GTGGGGAATT TATACCAAC ATCCATTGTA
2551 GGGGAATCT ATGATTCTGC TTCCCCAGCG GATTCCCACT CTGTCCACCA
2601 AGTGGGGGGT AGCACAGCCT CACAGCAACC GCCCTGACCT TGGGCAGTCT
2651 AGTGTTCCCTG CATTCTAGTC CCTGCTGTGC TGCAGGACTT TGGGCAAGTG
2701 ACCTGCCCTC TGTGAGCCTC CCTCTGACAC AGAGGAGGTG GCTCCCTTTC
2751 CCCACACCTT AGAGTGGCTG GGAGGGTAAC AAAGAGGGCC TGCCCCTTA
2801 GTCTCCTGCA CCCCTGCCCC CTGGTTTACC AGAGGGAGCG GATGAAGGAT
2851 GGCAGCATCT CACATGCCCC ATCACCACCT CTGAGGCACC TGGGGTGGGG
2901 GGGCGGAGCC CAGGCCTCTG GCTGCTCCCC TGTGGGAGCC ATTGGAATGT
2951 ATCCCCTGAC AGGCCCTT CCGCCTCCAC CTCAACCCAG GTCTTGGATT
3001 TCAGGTCCCT CCACCCCAT TCTGAGTCTC TGTCTTCTC CTTCCACCCG
3051 CTCCCAGGGT TTCCCACCAC AGGGTCTGGA AGTGTGTGTG ACGCCCATTG
3101 AGCTGTTACC CGAAGTCAGA TTAAAAATCA GGGAGTGTTT TCCCTCGTTT

```

FIGURE 1A

3151 CTGTAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AA (SEQ ID NO:1)

FEATURES:

5'UTR: 1-327
Start Codon: 328
Stop Codon: 937
3'UTR: 940

Homologous proteins:

Top 10 BLAST Hits

			Score	E
CRA 18000005055410	/altid=gi 5454030 /def=ref NP_006468.1	RAS-...	199	1e-49
CRA 18000004990267	/altid=gi 7438395 /def=pir T15833	hypotheti...	89	3e-16
CRA 18000005229906	/altid=gi 4884048 /def=emb CAB43324.1	(AL05...	86	2e-15
CRA 18000005189812	/altid=gi 9297040 /def=sp O94363	RHEB_SCHPO ...	85	3e-15
CRA 89000000193051	/altid=gi 7290026 /def=gb AAF45493.1	(AE003...	84	1e-14
CRA 89000000198900	/altid=gi 7296750 /def=gb AAF52029.1	(AE003...	83	2e-14
CRA 87000001027207	/altid=gi 7323471 /def=gb AAF59545.1	(AC024...	82	4e-14
CRA 40000057438157	/altid=gi 10119859 /def=dbj BAB13483.1	(AB0...	81	7e-14
CRA 18000004932656	/altid=gi 6981476 /def=ref NP_037348.1	Ras ...	81	9e-14
CRA 18000004905102	/altid=gi 5032041 /def=ref NP_005605.1	Ras ...	81	9e-14

BLAST dbEST hits:

gi 10220573 /dataset=dbest /taxon=96...	444	e-122
---	-----	-------

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

Expression information from BLAST dbEST hits:

gi|10220573 Lung small cell carcinoma

Expression information from PCR-based tissue screening panels:

Whole brain

FIGURE 1B

1 MVSTYRVAVL GARGVGKSAI VRQFLYNEFS EVCVPTTARR LYLPAVVMNG
51 HVHDLQILDF PPISAFPVNT LQEWADTCCR GLRSVHAYIL VYDICCDFS
101 EYVKTIRQQI LETRVIGTSE TPIIIIVGNKR DLQGRVIPR WNVSHLVRKT
151 WKCGYVECSA KYNWHILLF SELLKSVGCA RCKHVHAALR FQGALRRNRC
201 AIM (SEQ ID NO:2)

FEATURES:

Functional domains and key regions:

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site

142-145 NVSH

[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site

Number of matches: 5

1	4-6 TYR
2	37-39 TAR
3	105-107 TIR
4	150-152 TWK
5	159-161 SAK

[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

70-73 TLQE

[4] PDOC00008 PS00008 MYRISTYL
N-myristoylation site

Number of matches: 2

1	14-19 GVGKSA
2	178-183 GCARCK

[5] PDOC00266 PS00294 PRENYLATION
Prenyl group binding site (CAAX box)

200-203 CAIM

[6] PDOC00017 PS00017 ATP_GTP_A
ATP/GTP-binding site motif A (P-loop)

11-18 GARGVGKS

FIGURE 2A

BLAST Alignment to Top Hit:

```
>CRA|18000005055410 /altid=gi|5454030 /def=ref|NP_006468.1|
      RAS-related on chromosome 22 [Homo sapiens] /org=Homo
      sapiens /taxon=9606 /dataset=nraa /length=203
      Length = 203
```

```
Score = 199 bits (501), Expect = 1e-49
Identities = 105/204 (51%), Positives = 134/204 (65%), Gaps = 1/204 (0%)
Frame = +1
```

```
Query: 328 MVSTYRVAVLGARGVGKSAIVRQFLYNEFSEVCVPTTARRLYLPAVVMNGHVHDLQILDF 507
      M + RVAVLGA GVGK+AI+RQFL+ ++ E PT RLY PAV+++G V+DL I D
Sbjct: 1 MGGSLRVAVLGAPGVGKTAIIRQFLFGDYPERHRPTDGPRLYRPAVLLDGAVIDLSIRDG 60
```

```
Query: 508 PPIS-AFPVNTLQEWADTCCRGLRSVHAYILVYDICCDFSFEYVKTIQQILETRVIGTS 684
      +EW D L+ A++LVYDIC DSF+YVK +RQ+I ETR G
Sbjct: 61 DVAGPGSSPGGPPEWPDADKWSLQDTDAFVLVYDICSDFDYVKALRQRIAE TRPAGAP 120
```

```
Query: 685 ETPIIIVGNKRD LQRGRVIPRWNVSHLVKRKTWKCGYVECSAKYNWHILLFSELLKSVG C 864
      E PI++VGNKRD QR R PR ++ LVR+ W+CGY+ECSAKYNWH+L LF ELL+
Sbjct: 121 EAPILVVG NKDRQRLRFGPRRALAALVRRGWRCGYLECSAKYNWHVLR LRFRELLR-CAL 179
```

```
Query: 865 ARCKHVHAALRFQ GALRRNRCAIM 936
      R + H ALR QGAL RC++M
Sbjct: 180 VRARPAHPALRLQGALHPARCSLM 203 (SEQ ID NO:4)
```

Hammer search results (Pfam):

Model	Description	Score	E-value	N
PF00071	Ras family	44.3	4.9e-12	2
CE00060	CE00060 rab_ras_like	11.1	0.054	1
PF01118	Semialdehyde dehydrogenase	4.4	1.4	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF01118	1/1	5	14 ..	1	10 [.	4.4	1.4
PF00071	1/2	6	36 ..	1	31 [.	22.2	1.7e-05
CE00060	1/1	87	138 ..	99	150 ..	11.1	0.054
PF00071	2/2	89	161 ..	76	157 ..	22.1	1.8e-05

FIGURE 2B

```

1  GCGAACGTG GTTGAACAA ACAGCTGGCA GACTTGTGAC CCGGCCCTCG
51 GGATCCGCGA AGCCCCCGCT CTCAGCCTTG GGCAGCGACC CAGGTGTCCA
101 GACGCAGGGA AGGGGACGGA ACCAGGCTTC GCCCCCTGTG TGTGTCTCTG
151 GTCTCTTGCC TCTCTTTCCG GGCAGTCTCT TGCCGGGCGC TGTCTGCAGA
201 CTCACTGCAG AGCGCAGGCC TTGGGGAGAA AGCGCTCAGG GGCCTGGGCC
251 CTGCTTCCTG GGACAGCCCC CTCCCTTCCG CACTCAGCCA AGGGTGTGG
301 ATTAATATAC TCTTAGATCC AGGACCTCTG CCTGGAAAGA AGGAAGGGGC
351 ACCGATCGCT TCTGAGTTGG GGACAGGGCC ACACTTGGAC CTGATGAGTT
401 CCAGGAGGCT GGCCCCAGCC TAGGGGCCCT CTGCACCCCC TTCCTAGCTC
451 CTGGGTAGTG CCCCTCTTGC ATTTCCCTGC CGCTCCCCAG AAAGGGCTGG
501 TTTCTGGGCA GGGAGTGTGG TGTCCGCGAC AATCTGCAGG TAGGGTGCAT
551 CCTAGAAATC TCTGAGCTGC CACACCCAGA GAGAGGCCAG GACTCTTCTT
601 GCGGTTCCCT GCCCTCTCTT CCTCCCTTAG GACTCCCCAC TCCCCACCTG
651 CCTGTACTGG TTATGTAATT AACCAGCTG AGTCTCTCTG CCATGCTGGT
701 GGTAGTGGTG TCCAGGGATG CCTAAGGGGC TTCGGGACCT GGAGGGACAC
751 ATGGGGAGGA GGGATAGGTA CTTCCCCCTA GTTGGGAGCC CATGTAAGTG
801 TATAAACACC TGTGGAGAA GGGACACTGC ATGGGGCAGG AAGGAGTTCA
851 GGGTCCTAAT CTTAGTTAAC AACTCAACTG GCTATGAGGT CTTCACTTCC
901 CTAAGTCACT GGGTTTGT TTGGGGGGT TTGTTTGT TTAAACCTT
951 CAATAAAATG AGAAGAATCT CTCTCACGCA CACTGCATGC ATGTGAGCTC
1001 ATGCAACAGT AGATAGCAGC CCCTCCTGGG ATTGCTAGGA GGTTCCACTG
1051 CTAAGTTTGT GGAAGCTTA GATGAATGAA TGACCCCTGC CCATAGCAGT
1101 GGTCACTGTT TATTGAACAC CTACTCTGGC CTAGGCACTG GGCTTTACAT
1151 GCACACATCA TAACACTTAA TAAAGTTTCT TGCCCAAGGT CACACTTCAG
1201 GTGAATGGCA GAGCTAGGAA TCAAGCCCTG GCGGTTCTCA TTCCACATCC
1251 TGACCTGTAT TACTATGCT GTGTGCGCAC ATACCCCGTG GAGTGAGTCC
1301 TCTGAACCAG GACAACCTGG GGGACATTCA GCTGTGGCTT GTGCATATGT
1351 GGAAAGAACA TGCATCTGCA GAGAAGTATG CACCCATGCG GAAAAGCATG
1401 GCCTTGCCAG AATCCGGCGA ACCCACCTGC CCTGGAGCCA GTTAGCAAGT
1451 GCTGCTCTA CCCTCCTGGG AATCCAGACA GGAGGTCTCA GGCATGTAAC
1501 AGCCTCTCTG GTCACCATCA CCAACAAAAG AAGAAAGGTT CTCTCCTTTT
1551 CCTGGCTAAG AGATTTACTT GGATTTTCTC AGAAGTAGGG GCTGTTGACC
1601 TCATTTTACC TATGTGCAGA AGCATGGCTC CAGTTGAGAA GGTGATTTGC
1651 CCACTGACCT GCAGCAAGGA CACACCAGAG CTGATGAATT CGTATCAAGG
1701 CCAAACCCC TAGCAGCCCT TTTCTGGGCA CCTGCTATGT GCCAGGCCCT
1751 CTGGGCTCTA GGGGATGCCA GAGAATCAGA CCCAGACCTT GGCCTATATG
1801 CATTCACTGA TTAGAATCAC TGAACTTGGC CAGGCGCGGT GACTCACGCC
1851 TGTAATCCCA GCACTTTGGG AGGCCAAGGT GGGTAGATCA CCTGAGGTCA
1901 GGGATTCCAG GCCAACATGG CGAAACCCCA TCTCTACTAA AAATACCAA
1951 ATTATCTGGG TGTGGTGGCA GGTGCCTGTA ATCCCAGCTA CTCGGGAGGC
2001 TGAGGCAGGA AAATGGCTTG AACCAGGAG GTGAAGGCTG CAGTGAGTCG
2051 AGATCATGCC ACTGTACTCC AGCCAACTTG AAAGTTGGAA GAAGTACTGC
2101 GACATCGTAG AAAGGAGTAG CTGCCCCTCG GGTGTGATT GGGCACACTT
2151 CTTACTCCTT CTAAGCTTCT CTCTAAAGTG GGAATAACAG TACTTATCCC
2201 TTAGAGATGT CGTACTGACC AAGGATAACC ATATGGGCAT GCCAGCAGAG
2251 TAAGTGCTCA GCAGTGTTGC ACCCCAGAGG CGACGTGGTC ATCTAGTGCA
2301 GCATTCTCAA CTAGGGGCAT TTGGAAATGG GGGCAGGGGG AGTTTTTGGT
2351 CATCATGGTG TCTGGAGGCT GGCAGTTACC TCTTGGGGGC CACGGAAAGC
2401 AAATGTCTTG CAATGTGTGG GCAGTCTTGC ACAATCGAGT TCTCTCCCA
2451 CAAAATGTCA GGAGTGCCAG CACTGAGACA CACCTGCCCA GTCCCACCA
2501 TTCAGGAGGA CACAGACTCA GAGGTGTTGC CGTCTTGTCC CAGGCTCTGT
2551 GGGGAAGCTG GGATCAAACC AAGTCCAGTG CGCTTCCCAC TCTGCTCTGC
2601 AGCCTGTTTT GGTGGAGTT GGACCTGGAG AAAAGTCAAG TCATAAGTCA
2651 AGAAAGATTG GGCCCTACTA CTGGAATGCA GGAAAAAATG GAGGAGGGAT
2701 GGAGAGGTTT TGGAAAGGCA GCCACAGGGG TTCTGGGAGA GGAAGGCAT
2751 TCTAAGTGCC AGTAACAGCT TCAGCAAAGT CCAAAGGTG GAAAAGTGCA
2801 GGACACGTCC AGGATAAGC CAGTGCAC TAAGCCACCTC TTGTCCCCAC
2851 AGTCCAGGTG GAGGCCGAG AGGGCCCAGG GCAAGCAGAG GCAGCAATGG
2901 TTGGTCTCTG CGGTGGCTGA GCCCCAGCC CCTGGAATAT GCAGCCCGGG
2951 GGAGCCCCAG ACAGCGGCAA GGACGAGGTG GCGGAGTGGG GCGGGAGGCA
3001 TGGTCTCCAC CTACCGGGTG GCCGTGCTGG GGGCGCGAGG TGTGGGCAAG
3051 AGTGCCATCG TGCGCCAGTT CTTGTACAAC GAGTTCAGCG AGGTCTGCGT
3101 CCCCACCACC GCCCGCCGCC TTTACCTGCC TGCTGTGCTC ATGAACGGCC

```

FIGURE 3A

3151 ACGTGCACGA CCTCCAGATC CTCGACTTTC CACCCATCAG CGCCTTCCCT
3201 GTCAATACGC TCCAGGTAGG AGGACCCTGG GGGGCATGGG TTAGTGGGGA
3251 AACGGATGGG TAGGGGAGAG GCTGGATTCC AAAC TGCTGT AGCTTGGGCC
3301 CTATTGCCAG GGGCCCATCA CTGAGTTTGG GAGCTCCACA CTGCACCTTG
3351 GGCCACTCTG CTTAGAGCCG TTCCAGGAAT CCATTCAATTG GTGTGCTAGT
3401 TTATTCAACA AATATTTGGT GACCGTTCAA TGTGTGCCAG GCCCTGCAGT
3451 GGGCACTGGT GCAGAATGGT GAGCAAAAAA TATATGGAAT TTGCTTTCAA
3501 GAAACTCATA GTCTGGTGAG AAAAGGCAAA TATGGTGTGA TAAGTTCTAT
3551 GATTGGAGGA GCAGGGAGCT GGGGCAGCCC TTAAGGGGGC ATCTAGGCCA
3601 TCCAGATGTG TTGGGGTGGA GTTGGGGGGT CACAGAGGGT GATGTCTCAA
3651 CTAAATAGGT TTTAGGCAGG TAAGAGTCAG TAGAGAAAAG GACAGGGAAC
3701 ACTAGGCTAC TGTGAGTATT CGGAGCTGTG CCTACCGTAA CCTCACTCCA
3751 CATCCTCTGG AGAAGGGACA GCAGCAGAAC AGACGGGGCC CTGGGAAAGG
3801 TGTGTTCTTG GAGACTCTGG AGACCCAGT CAGGTCTCTT GCCCAAGGCC
3851 CTCTTCTCTT AAGTGATGCT CTGCCCCTGA CCTCAGGACC TGCCTGCTGG
3901 GCACCCTCCC TGCCAGGTTT GGATTTAAAT GCCTGAGGGT CCTCACTTAT
3951 TGTGTTCCCT CCCCACTGCC TGCTGGAACC AGGTCTCTCT GCCCTCTCTC
4001 AACCTCTGAC TTGAGAGGGA GTGGAGAGAA AAAGGAAAGT GAGCTCTAGG
4051 ACATGTTTGC TCACTGAAGG AAGCCTCTGA CCAGAGTGTA CAGAGCTTTT
4101 CCAGGAAGGA CAGGCACAGT GGTGGAGGCC CAGAAGACAG GGGACAAGGC
4151 TCGTCCAGGT GTAAC TGAGC AAATCAAGCA GTCTCTCAGG CTGAGACCCT
4201 GGGCTGGGAG ATGGCGGGCA GCTCAGCACT CAGCACTCTC GGCAACACCA
4251 GGCAGGAGGG CCCTGCCCTA ATCTGCCGGA GACACCTGTT CACCCATCCC
4301 AGGCACCTGG GGTCAAGGAG AAAGATGGAA GCCTGATCCC GCATCTGCCC
4351 TGGAAGCAGT GAGGCTGAGC CTGTCAGGGC AGACAGTCTG GATGCAGGGC
4401 CTTCTAGTTC TCTTCTAAAG GAGACTTTAA CAATCACCTG ATTGGACATT
4451 CAAATCTTGC TCCAAGCCTA CACACTGAGC TTTGTTGATT TCATCTTGCC
4501 CCCTTTTACCT TGATTCTCTG CCCACTCTCT ATAACCCTC TTATCGAATT
4551 TTTCTTTCTT TTTTAAAATT TATTTATTTT TTTATTTTAG ATGGAGTCTC
4601 CCTCTGTCGC CCAGCTGGA GTGCAGTGGC ACGATCTCGG CTCACTGCAA
4651 TCTTCGCCTC CCGGGTTCAA GCGATTCTCC TGCCTCAGCC TCCTGAGTAG
4701 CTGGATTACA GGCACCTGCG ACCACACCCA GCTAATTTT GTATTTT TAG
4751 TAGAGATGGG GTTTCACCAT GTTGGCCAGG CTGGTCTCAA ACTCCTGACC
4801 TCAAGTGATC CGCCTGCCTA GGCCTCCCAA AGTACTGGGA TTATAGGCAT
4851 GAGCCACCAC GCCTGGTCTC TTATCCATAC TTTCAGTGTT TCTTTACCCA
4901 AGTAAGAAAA TGCATTCTTC CCTGCTTCTT ACGTAAAGAA CAAAACAAAA
4951 ACAAGAACCA TACTGTTCTG TACCTTGATT TTATTTTATT TTTAAAATT
5001 TTTGTATAGA TGGGTCTTGC TGTGTTACCT AAGCTGATCT CGAACTACTG
5051 GCCTCAAGCG ATCCTCCTGC TTTGGCCTCT CAAAGTGCTG GGATTACAAG
5101 TGTGAGCCAC TGTGCTTGGC GCTGTACCTC AATTTTTTTA ACTTGCTATT
5151 ATAACCTGAA GATTTTTTCCA GGCCATTATC TAGAGGACGT CCTCATCTT
5201 TTTTCATGGC CACGCCCTAC TCCATTGAAG AGCTATACCA TGGAGTCCCT
5251 TCTTGTTGGA TAAGTGGGTG GTATCCAGTC TTGTGCTGTT TCAAACAGTG
5301 CCAACAATGAG TGGCCTTGTA GATAGGTCAT TTTGAACATA AGTAGGTATA
5351 TCTGTGGGAT CAATTACCGG AAAGGGCATT GCTGGAAATG GCACTGCTGG
5401 ATCACAATGC CTGGAAATGG CATTGTGAAT ACAGAGCCAG GTGAGGTGGC
5451 TCATGCCTAT AATCCCAACA CTTTGAGAGG CTGAGGCAGG CGGATCACTT
5501 GAGCTCAGGA GTTCGAGACC AGCCTGGGCA ACATGACAAA ACTCCGTCTC
5551 TACCAAAAAT AAAAAAATT AGCCAGGCAT GGTGCTCCAT GCCTGTGGTC
5601 CCAGCTGCTT GGGAGGCTGA GGTGGGAGAA TCGCCTGAGG CCGGAGGTT
5651 GAGGCTGCAG TGAGCTGAGA GTGCCACTGC ACTCCAACCT GGGTGACAGA
5701 GTGAGGCCCT GTCTCAAAAA AAAAAAAAAA AAAGTGTGAC TGTAAC TGGA
5751 GTTTGGAGGG GAGGTTATTT CCAGATTGCC CTCCATAGCA GTGGCGTATG
5801 CTGTGCTCCT GTGAGCAATG TATATGAGAG CCCGTTTTCC TACAGTCTTG
5851 CCATCAGAGT ATATTGTCAA ACTTTTGACA ATATATTTGA CAATCTGAGA
5901 GATGAGATAT GATATTCTCC TTGTAGTCTC CATTTGCATC TCTGATCGTG
5951 GGTGAAATTG AGCATCTTTC ATAGGTTTAA GGGCCTTTGT GTTCTCTTT
6001 TCAAGAACTA TTGATGTCTT TTGCCCATTT TTCTATTGGG TTGTTGGCTT
6051 TTTTCTTCTT GACTGACCTT GAGTTTTGGA CTCTAAGATA TCCAAGATT
6101 CACTCCTGGA GCCCAGTAAG GGACTTTTGG CAGAGAAAATA CTGTGAAAAA
6151 GGTATCCTCA AGGCACCAAA GATTAAGTAT AAAACCTAAG AATCCTGATG
6201 GCCACCATCT GGAAACAAAA TAATACATTC TTCTCCAATG CCAGATGAGA
6251 TAGAGCCCAG GAGAGTAGTG TTTCTGGGT GTGAGCCTCA GTGTCTTCTG

FIGURE 3B

6301	CAGCCCCCTTC	TATCAGAGAA	GGAAGCTGAG	ATTATCAGGT	GCTTGCAACT
6351	CACCAAAGGA	ATTATCAGCA	AATGCATGGT	TGAGATGCAG	GTGGCTGAGC
6401	CTTGTCCTTG	AAACTGGACT	CCCTTTCTAT	TGCTCCTTCT	CTGTCTTGAC
6451	AGAGCCCCAA	GATGGCCCTT	TACAGTTTGG	AACCCCTGCTT	CCTCCCTTCA
6501	ATCAAGGGGG	AAGGGATAAG	CTAGCCAATC	AGGGGCCCTC	CTCCTCTCTC
6551	TTTTAGGAAC	CCCCAGAGAG	GAGTGGGTGG	GAGGAAGCCA	GGAGTTCCCC
6601	TCAAGGAGGC	AACATGTTGG	GGGAGAGGTG	GGGCTGTAC	CCTCAAAAGC
6651	TGGCAGCTGC	TCCCTCTCCC	CAGCAGACAG	CTTGAAGAGA	CTGGGAGCTT
6701	CTCATCCCTC	CCACTTCTCA	CTGATCTCCA	TTGGTCTTGG	GGGATCGTGG
6751	GAGCATCCGT	ATACACAGGT	TCCAGGCTCC	TGGAGATCAC	TGTGTCCAGC
6801	AGAATGCACT	CTTCCCTGGC	CTAAGAAACC	AGTTTCCCTAT	GGTTTTAGGT
6851	TTGTCTCTCG	CATCCTCCCG	CCGCACCAAA	AATTTAAACC	TCAGCACAAA
6901	GAAAAGATGC	CACATCATCT	CCCTAGGGAA	ATCCACTGCA	GCATCTTCTA
6951	AGCCTTTTGG	TTGGGAAGTG	CTGTTCTGAA	GTTGGACTTA	ACTCTGCACT
7001	ACTGCCACCA	AAGTCGTTTC	CTTTTGATCC	TTCTTGGAAG	TGGAGAACTG
7051	TAGTCCTCCT	TTGTGCCTGG	CCCCTGCCCC	ACTCAATTCA	GATGCTGGGA
7101	CAGGAGACAT	ACCTCCACCT	TCTTCTAGTC	TTTTGCCTGG	GCTTTGGTGG
7151	GAGAAGACTC	TGGTTTCCCT	TGTCCTTGGA	GGCCTCTGTC	CCCCCACCTT
7201	TAGGGACCCC	CTTCTTTCCA	CACACTGGCT	GCCTGAAACC	GCTCTTGCAG
7251	CTGGCAGGTT	GACTAATGAA	TCTGTTAAGG	AAACTTCTCT	TTAGTGTACT
7301	TGGCCTTTCT	AGGAGTCTCT	TCACCTTGAG	CTGTACCCCC	CAATCCCTTG
7351	AGAAGTTGCC	ACAAACATTC	AGGAAGTTCA	TCTCCCTGGA	GCTGCCCAGG
7401	GGCCCTACTC	TACATCAGCC	CATTATGCAT	CCAGTCTGAA	TCTTTTTCTG
7451	TTTCTCAACC	CTGAGGGCAG	AGAGAAGCAT	ACAGAAGGGG	CACATCAGGT
7501	AGCAGTCTAA	GGGCAGTGGC	AGAGGCAGGA	GTTGCATTGA	TCCCAGCTTG
7551	GGCCATGGAG	AGCTCACCAG	CCCAGGTAGT	GCTATTAAGG	AGCACCTGCT
7601	TTGAGCCAAC	AGTGCTAGAC	ACTCAGGGAG	GAAGAGGGAG	TATATACAAA
7651	TGAGGATGGC	CTGGCTGTGG	CCTTCTCAGG	AGCTCACAGC	AGAAGTGGGG
7701	AACTGGAGAT	GGAACAGCTC	TAATGAAAGT	GTAATAGAAG	GATTGTTAGA
7751	ACACAGTGGC	TGAAGGGAGT	AGTCCCCGCT	TTGCAGGAAG	GATGGGAAAT
7801	CAGGAAGCTT	CTTGGAGGTG	GTGGCACTTT	AGCTGAACCT	TGGAAGATAG
7851	AATTTTAACA	GGTCCAACAC	CCAGCTCAGA	GCTGGACTCT	TAGAGGTACT
7901	TAATAAATGT	ACTTGTTGAA	CAAAGGCCTC	GATGGATGGA	TGAGGGCACG
7951	ACATGGAGCA	AGGCAGAGCT	AAACTCCAGA	TGTGCACAAG	ACAGTGCAGT
8001	GGCCCTGTAG	ATCAAACAAT	GTGACCTGCT	CCATCCTGGC	TTGGGAATGG
8051	GGAGGCTACA	GCTCCTCCAT	TCTCCCTGGG	CCTGGTCTCC	TGGGGATGGT
8101	CGGGTATGGA	AGGCTTCAGG	TGCAGTGGCA	GGTGAGAGCA	CTGCCCTCT
8151	GATGGGAGGT	GTTTGGGGGC	TAGGGGAGCC	CTCATGGCTG	CTCTGACCCT
8201	GGTACTGGCT	GGGGATATTG	CAGGAGTGGG	CAGACACCTG	CTGCAGGGGA
8251	CTCCGGAGTG	TCCACGCCTA	CATCCTGGTC	TACGACATCT	GCTGCTTTGA
8301	CAGCTTTTGG	TACGTCAAGA	CCATCCGCCA	GCAGATCCTG	GAGACGAGGT
8351	GAGAGGCTGG	AACACAGTCC	ATTGCCACCT	CTGTGGATGC	CCCAGTGCTA
8401	GCCAGTCCCT	GTGAAAAGGG	CACAGTATAG	GGACACAGAT	AGAGGTATAT
8451	GTGTTCTAAG	ATTTCCACAC	ATACACTCAA	ACATGCATAC	ATTGTGCTGT
8501	TCCCATTCTC	GTCAACTCAT	GTTGGGACCG	TGGCTGTGGG	GGTGGCTAGA
8551	GTAGTGCAGT	AGTTAAGAAC	TGGGACTTCT	GGAACAAGAC	TTCCAGGGCC
8601	ACTCAGCTGC	ATGACTTGAA	GCCAGTAAAC	ATTTAAGCCT	ATGTCCTCAT
8651	CTGTAAAATG	GGGATAACAG	TAGAACCCAT	CTTTTAGATC	AGTTGTGCTG
8701	ATCAGAGAAT	ATAACACCTC	CAGGGCTTAG	GGCTGCGCCT	GGAGCAGAAC
8751	CTACGGTGGT	GGTAGTATTG	GCCAGGCACA	GCCTGCCCTG	CTGGGAGTAC
8801	AGCGGTTGTG	GGGCTGACAG	AGTTCTGAGC	TGCCTGCCTC	GCCCCACAGG
8851	GTGATCGGAA	CCTCAGAGAC	GCCCATCATC	ATCGTGGGCA	ACAAGCGGGA
8901	CCTGCAGCGC	GGACGCGTGA	TCCCGCGCTG	GAACGTGTGC	CACCTGGTAC
8951	GCAAGACCTG	GAAGTGCGGC	TACGTGGAAT	GCTCGGCCAA	GTACAACCTG
9001	CACATCCTGC	TGCTCTTCAG	CGAGCTGCTC	AAGAGCGTCG	GCTGCGCCCC
9051	TTGCAAGCAC	GTGCACGCTG	CCCTGCGCTT	CCAGGGCGCG	CTGCGCCGCA
9101	ACCGCTGCGC	CATCATGTGA	CGCCTGCGCG	CCCCTCGGGC	TGCACCCGCA
9151	CTGGCCGAGC	GGAGGGCGGG	GCCGTACTGC	GGGGCTGGGG	CGGGGAGCGG
9201	GCGGGAAATG	GAAGTGTGAC	GGTCCCGGCC	TGAGGCCCCC	GCAGCCACGC
9251	ACCTCCCGGT	GAGAAGCAGA	GCGCGAGAGG	GAGCCCTCCG	TAACTGCCCA
9301	GCCCTGCCCC	TTGCCCCCGT	GGCTTCTCTG	GACAGCCGCC	TTCAGTGCCTG
9351	TATTTAGTGC	AGTGCCCGGC	CCGACCCGCG	GGGGTGCCAC	AGCCTTTTGG
9401	GATGGGGGTG	AGCGTGCAAT	GGAGGCTGGG	GGTGCCGAGG	TGCCGCCCTG

FIGURE 3C

9451 GCCGGGCCCC CACGTGTCTT CTCCAGAATG TGTCTGTCTT TGCCCTGGTGT
 9501 CTTCTTTTCC CGTGTCCGCC CACCCAGCG TCTGTTGGTA CTTACCTGTC
 9551 TCACCTACCC TCAGTCCCC TCCCAGCTCC GCTCACAGGG CTCTCATTTT
 9601 GTCCATCCCC TTGTCGCAGA TCCTGGCAGC TTCTTTGTGA GGCCAGGCCT
 9651 TCTGACTGTC AGCACCACCG GCACAGGGCA GAGATGCGGG TGGCCCAAGG
 9701 ACCACGATCA AGGGGTCCGG GGGACCGAGG TCCCAGATCA GTGAGGGGAG
 9751 AAGGTTGAGC TCTCCGGCTT CCAGGGAGAC CTCCCCGCCC AGCAGCCCCC
 9801 AGAGACACAA CAACCTACCT TCCAGCCTTA ACTCGATGGT CCGTCCCTGC
 9851 CAGGTGCCCC TCACTCTTCC TGACCCCAAA GCCAGATCAC CCCCTGGGTT
 9901 AAAACTTTTT TTCTTTTTTT TTTTGTGACA GAGTGTGGAA AGGGAGCCCC
 9951 CCAAAGGATA GCTTCTTTTT CATGATGCCA GGCTCCAGTC CTTTATTCCC
 10001 TTCTGCATAC TGCAATCTGA TCTGTCAGAC TGGGGAATGT TGGGTCTCTGG
 10051 GGTCTGGTCG TGGGCAGGAT GGTGCCCAGA AGGGGGTAG GTTGTCCCAG
 10101 TGAAAATTCT GTTGCCCCGT CTCAACCCCA TCTGACTACC CCAGACTCTG
 10151 CCTGCCTCAG ATCTCAGACT ATCCTGATTA ATCTGGGGAA GAACAGAGCC
 10201 AGGGAAGAA TGGTGGGGAC CCCTGTACTT GGGGGAGACA CACCTGCATC
 10251 TTCTCTCGT CACAGATGGA GGCCCTCAGG ATCTGACACC CTCTGTGCC
 10301 AACACCAGTC AGCCCTATAC CCTAACTCAC TCCACCCCAT TTTCTCCGGC
 10351 TGCCTGGCCG GGTTCCTACC TCTCGTCACC GGAGCTGATC ACTGTCAGTT
 10401 TTGTACCGAT TTAGAAATA CAATAATAAT GAAGATTCTA GGAATGGCAT
 10451 GAGGGATTGA TGGGGGACTT GGAGGGAGGG ACAAGTGGTG CCCTGTCCCC
 10501 TGCTCCCCTG GCCAAAGAAA GCTGTCCCTG AGGCTGAGCC CTCAGCCCTG
 10551 GCCTGGTGGG GGGACAGCAA GGTCCCTTGT TATAAGAGGG GCAGAGAGGA
 10601 CAACTCCGCT TTGGCCAACC TAGCCAAGGC TGCAGCATAT AGACCAGGAA
 10651 ATCAGGTAGC CCAGACTGGT GATGGAGCAG AGTCTGGGGG AAGGGTCCTG
 10701 GGTGGGGAAT TTATACCAA CATCCATTGT AGGGGGAATC TATGATTCTG
 10751 CTTCCCCAGC GGATTCCAC TCTGTCCACC AAGTGGGGGG TAGCACAGCC
 10801 TCACAGCAAC CGCCCTGACC TTGGGCAGTC TAGTGTTCCT GCATTCTAGT
 10851 CCCTCTGTG CTGCAGGACT TTGGGCAAGT GACCTGCCCT CTGTGAGCCT
 10901 CCCTCTGACA CAGAGGAGGT GGCTCCCTT CCCCACACCT TAGAGTGGCT
 10951 GGGAGGGTAA CAAAGAGGGC CTGCCCCCTT AGTCTCCTGC ACCCCTGCCC
 11001 CCTGGTTTCA CAGAGGGAGC GGATGAAGGA TGGCAGCATC TCACATGCCC
 11051 CATCACCAAC TCTGAGGCAC CTGGGGTGGG GGGGCGGAGC CCAGGCCTCT
 11101 GGCTGCTCCC CTGTGGGAGC CATTGGAATG TATCCCCTGA CAGGCCCCCT
 11151 TCCGCCTCCA CCTCAACCCA GGTCTTGGAT TTCAGTCCC TCCACCCCA
 11201 TTCTGAGTCT CTGTCTTCT CTTCCACCC GCTCCCAGGG TTTCCACCA
 11251 CAGGGTCTGG AAGTGTGTGT GACGCCCATT GAGCTGTTAC CCGAAGTCAG
 11301 ATTAATAATC AGGGAGTGTT TTCCCTCGTT TCTGTACCAA GGTGTTGGCT
 11351 CCATTCTCTA TGGTAGGAGG GGAGGGGTCC CCACAGGGCT TGCCTGCTGA
 11401 GCTCCGTGTG GAAGGAGGGT GAAGGTGGTG AGGTGGCCCC CAGTCCCAA
 11451 GCCCAGGTCA ACAGGGAGAC CACCGGTGAA GAGTTTGGGA TTTATCACCT
 11501 TTCCACCTAA CCCCAAACCC TCCAGCTAAT TCCAACCAT CAGAAGGGAA
 11551 GCAGAACTTC TCCCCTGCCA CTGTCTGGAA AATTTCATA ATGGGACTCA
 11601 ATCCCAGCTT TCCCGTCTGC GTCTCGTCCT TCCCACCTAA GGCTGAGACT
 11651 TTACAGCCTC TCAGTCATAA CTTCTTGGAT GTAGATGTGT TAGGAACACT
 11701 TTCAGCCACC CGTCTTGTCC CTGAGTGATC TCAGGTCCCA AACTCCAGAG
 11751 CAAAGCTTTG AAATCTTGGG CAAGGGTGCC TTGTGGGAGC CTGTGTGTTG
 11801 AGGGCAGGAC TGGTCTCTGT CCGTGGTGCT GACCCACCAG CCACTTCCAG
 11851 GAAAGATGGG GCTGCCGTC AAGGTTGGCT GAGCCTCAA AGAGGAAGCC
 11901 TCTCTACCA CCAACTCCTT CCTTCTAGTC CCCATCTCCT CCAGTGGGAT
 11951 AACATCTGAA GCTATACCTC CCCGCACCAC CACAGTCCTG GAGTGAGGGA
 12001 CTCAAGAAGC TGGGGGGCAG GGGGAGGCAG GTTCAGTGGT TCACATCTTT
 12051 AATCCCACTG CTTTGGGAGG CCAAGGCAGG AGGATCGCTT GAGGCCAGCC
 12101 TGGACAACAT AGTAAGAC (SEQ ID NO:3)

FEATURES:

Start: 3000
 Exon: 3000-3215
 Intron: 3216-8223
 Exon: 8224-8348
 Intron: 8349-8849
 Exon: 8850-9117
 Stop: 9118

FIGURE 3D

CHROMOSOME MAP POSITION:

Chromosome 17

ALLELIC VARIANTS (SNPs):

DNA

Position	Major	Minor	Domain
2455	A	C	Beyond ORF (5')
2785	A	G	Beyond ORF (5')
3482	T	A	Intron
6189	A	G	Intron
6491	T	C	Intron
7353	A	T	Intron
8688	A	G	Intron
10789	G	C	Beyond ORF (3')
11079	G	A	Beyond ORF (3')
12087	A	G	Beyond ORF (3')

Context:

DNA

Position

2455	CTCCTTCTAAGCTTCTCTCTAAAGTGGGAATAACAGTACTTATCCCTTAGAGATGTCGTA CTGACCAAGGATAACCATATGGGCATGCCAGCAGAGTAAGTGCTCAGCAGTGTGTCACCC CAGAGGCGACGTGGTCATCTAGTGCAGCATTCTCAACTAGGGGCATTGGAATGGGGC AGGGGAGTTTTTGGTCATCATGGTGTCTGGAGGCTGGCAGTTACCTCTTGGGGGCCACG GAAAGCAAATGTCTCGCAATGTGTGGGCAGTCTGCACAATCGAGTTCTCTCCCCACAAA [A, C] TGTCAGGAGTGCCAGCACTGAGACACACCTGCCCAGTCCCACCCATTCAGGAGGACACAG ACTCAGAGGTGTTGCCGTCTTGTCCCAGGCTCTGTGGGAAGCTGGGATCAAACCAAGTC CAGTGCCTTCCCACCTCTGCTCTGCAGCCTGTTTTGGTTGGAGTTGGACCTGGAGAAAAG TCAAGTCATAAGTCAAGAAAGATTGGGCCCTACTACTGGAATGCAGGAAAAAATGGAGGA GGGATGGAGAGGTTTTGGAAAGGCAGCCACAGGGTTCTGGGAGAGGGAAGGCATTCTAA
2785	TGCCCAGTCCCACCCATTCAGGAGGACACAGACTCAGAGGTGTTGCCGTCTTGTCCCAGG CTCTGTGGGGAAGCTGGGATCAAACCAAGTCCAGTGCCTTCCCACCTCTGCTCTGCAGCC TGTTTTGGTTGGAGTTGGACCTGGAGAAAAGTCAAGTCATAAGTCAAGAAAGATTGGGCC CTACTACTGGAATGCAGGAAAAAATGGAGGAGGGATGGAGAGGTTTTGGAAAGGCAGCCA CAGGGTTCTGGGAGAGGGAAGGCATTCTAAGTGGCAGTAACAGCTTCAGCAAAGTCCCA [A, G] AGGTGGAAAAGTGCAGGACACGTCCAGGGATAAGCCAGTGCCTAAGCCCACCTCTTGTC CCCACAGTCCAGGTGGAGGCCCGCAGAGGGCCAGGGCAAGCAGAGGCAGCAATGGTTGGT CCTGACGGTGGCTGAGCCCCCAGCCCCGGAATATGCAGCCCGGGGAGCCCCAGACAGC GGCAAGGACGAGGTGGCGGAGTGGGGCGGGAGGCATGGTCTCCACCTACCGGGTGGCCGT GCTGGGGCGCGAGGTGTGGGCAAGAGTGCCATCGTGCGCCAGTTCTTGTACAACGAGTT
3482	ACCCATCAGCGCCTTCCCTGTCAATACGCTCCAGGTAGGAGGACCTGGGGGGCATGGGT TAGTGGGGAACGGATGGGTAGGGGAGAGGCTGGATTCCAACTGCTGTAGCTTGGGCCC TATTGCCAGGGCCCCATCACTGAGTTTGGGAGCTCCACACTGCACCTTGGGCCACTCTGC TTAGAGCCGTTCCAGGAATCCATTTCATTGGTGTGCTAGTTTATTCAACAAATATTTGGTG ACCGTTCAATGTGTGCCAGGCCCTGCAGTGGGCACTGGTGCAGAATGGTGAGCAAAAAAT [T, A] TATGGAATTTGCTTTCAAGAACTCATAGTCTGGTGAGAAAAGGCAATATGGTGTGATA AGTTCTATGATTGGAGGAGCAGGGAGCTGGGGCAGCCCTTAAGGGGCGATCTAGGCCATC CAGATGTGTTGGGTGGAGTTGGGGGTCACAGAGGTGATGTCTCAACTAAATAGGTTT TAGGCAGGTAAGAGTCAGTAGAGAAAAGGACAGGGAACACTAGGCTACTGTGAGTATTTCG GAGCTGTGCCTACCGTAACCTCACCTCCACATCCTCTGGAGAAGGGACAGCAGCAGAACAG
6189	GACAATCTGAGAGATGAGATATGATATTCTCCTTGTAGTCTCCATTTGCATCTCTGATCG TGGGTGAAATTGAGCATCTTTTCATAGGTTTAAGGGCCTTTGTGTTTCTCTTTTCAAGAAC TATTGATGTCCTTTGCCCATTTTTCTATTGGGTGTTGGCTTTTTTCTCTTGACTGACC CTGAGTTTTGGACTCTAAGATATCCAAGATTTCACCTCGAGGCCAGTAAGGGACTTTT

FIGURE 3E

GGCAGAGAAATACTGTGAAAAAGGTATCCTCAAGGCACCAAAGATTAAGTATAAAACCTA
[A, G]
GAATCCTGATGGCCACCATCTGGAACAAAATAATACATTCTTCTCCAATGCCAGATGAG
ATAGAGCCCAGGAGAGTAGTGTTCCTGGGTGTGAGCCTCAGTGTCTTCTGCAGCCCTT
CTATCAGAGAAGGAAGCTGAGATTATCAGGTGCTTGCAACTCACCAAAGGAATTATCAGC
AAATGCATGGTTGAGATGCAGGTGGCTGAGCCTTGTCCTGAAACTGGACTCCCTTTCTA
TTGCTCCTTCTCTGTCTTGACAGAGCCCCAAGATGGCCTTTTACAGTTTGAACCTTGCT

6491 AATCCTGATGGCCACCATCTGGAACAAAATAATACATTCTTCTCCAATGCCAGATGAGA
TAGAGCCCAGGAGAGTAGTGTTCCTGGGTGTGAGCCTCAGTGTCTTCTGCAGCCCTT
TATCAGAGAAGGAAGCTGAGATTATCAGGTGCTTGCAACTCACCAAAGGAATTATCAGCA
AATGCATGGTTGAGATGCAGGTGGCTGAGCCTTGTCCTGAAACTGGACTCCCTTTCTAT
TGCTCCTTCTCTGTCTTGACAGAGCCCCAAGATGGCCTTTTACAGTTTGAACCTTGCTT
[T, C]
CTCCCTTCAATCAAGGGGAAGGGATAAGCTAGCCAATCAGGGGCCTTCTCCTCTCTCT
TTTAGGAACCCCCAGAGAGGAGTGGGTGGGAGGAAGCCAGGAGTTCCCTCAAGGAGGCC
ACATGTTGGGGGAGAGGTGGGCCTGTCACCCTCAAAGCTGGCAGCTGCTCCCTCTCCCT
AGCAGACAGCTTGAAGAGACTGGGAGCTTCTCATCCCTCCCACTTCTCACTGATCTCCAT
TGGTCTTGGGGGATCGTGGGAGCATCCGTATACACAGGTTCCAGGCTCCTGGAGATCACT

7353 GTCCTCCTTTGTGCTGGCCCCTGCCCCACTCAATTCAGATGCTGGGACAGGAGACATAC
CTCCACCTTCTTCTAGTCTTTTGCTGGGCTTTGGTGGGAGAAGACTCTGGTTTCTTTG
TCCTTGGAGGCCCTCTGTCCCCCACCCTTAGGGACCCCTTCTTTCCACACACTGGCTGC
CTGAAACCGCTCTTGCAGCTGGCAGCTTGACTAATGAATCTGTTAAGGAAACTTCTCTTT
AGTGTACTTGGCCTTCTAGGAGTCTCTTACCTTGAGCTGTACCCCCAATCCCTTGAG
[A, T]
AGTTGCCACAAACATTCAGGAAGTTTATCTCCCTGGAGCTGCCAGGGGCCCTACTCTAC
ATCAGCCCATTATGCATCCAGTCTGAATCTTTTCTGTTTCTCAACCTGAGGGCAGAGA
GAAGCATACAGAAGGGGCACATCAGGTAGCAGTCTAAGGGCAGTGGCAGAGGCAGGAGTT
GCATTGATCCCACTTGGGCCATGGAGAGCTCACCAGCCAGGTAGTGTCTATTAAGGAGC
ACCTGCTTTGAGCCAACAGTGTCTAGACACTCAGGGAGGAAGAGGGAGTATATACAAATGA

8688 TGCCCCAGTGCTAGCCAGTCCCTGTGAAAAGGGCACAGTATAGGGACACAGATAGAGGTA
TATGTGTTCTAAGATTTCCACACATACACTCAAACATGCATACATTGTGCTGTTCCCAT
TCTGTCAACTCATGTTGGGACCGTGGCTGTGGGGGTGGCTAGAGTAGTGCAGTAGTTAAG
AACTGGGACTTCTGGAACAAGACTTCCAGGGCCACTCAGCTGCATGACTTGAAGCCAGTA
AACATTTAAGCCTATGTCTCATCTGTAAAATGGGGATAACAGTAGAACCCATCTTTTAG
[A, G]
TCAGTTGTGCTGATCAGAGAATATAACACCTCCAGGGCTTAGGGCTGCGCCTGGAGCAGA
ACCTACGGTGGTGGTAGTATTGGCCAGGCACAGCCTGCCCTGCTGGGAGTACAGCGTTG
TGGGGCTGACAGAGTTCTGAGCTGCCTGCCTCGCCCCACAGGGTGAATCGGAACCTCAGAG
ACGCCCATCATCATCGTGGGCAACAAGCGGGACCTGCAGCGCGGACGCGTGATCCCGCGC
TGGAACGTGTGCGACCTGGTACGCAAGACCTGGAAGTGGCGCTACGTGGAATGCTCGGCC

10789 TGCCCTGTCCCTGCTCCCTTGGCCAAAGAAAGCTGTCTTGAGGCTGAGCCCTCAGCCC
TGGCCTGGTGGGGGACAGCAAGGTCCCTTGTTATAAGAGGGGACAGAGGACAACCTCCG
CTTTGGCCAACCTAGCCAAGGCTGCAGCATATAGACCAGGAAATCAGGTAGCCCACTG
GTGATGGAGCAGAGTCTGGGGGAAGGGTCGTGGGTGGGGAATTTATCACCAACATCCATT
GTAGGGGAATCTATGATTCTGCTTCCCCAGCGGATTCCCACTCTGTCCACCAAGTGGGG
[G, C]
GTAGCACAGCCTCAGCAACCGCCCTGACCTTGGGCAGTCTAGTGTTCCTGCATTCTAG
TCCCTGCTGTGCTGCAGGACTTTGGGCAAGTGACCTGCCCTCTGTGAGCCTCCCTCTGAC
ACAGAGGAGGTGGCTCCCTTCCCCACACCTTAGAGTGGCTGGGAGGGTAACAAAGAGGG
CTTGCCCCTTTAGTCTCCTGCACCCCTGCCCCCTGGTTTACCAGAGGGAGCGGATGAAGG
ATGGCAGCATCTACATGCCCCATCACCAACTCTGAGGCACCTGGGGTGGGGGGGCGGAG

11079 CCAAGTGGGGGGTAGCACAGCCTCAGCAACCGCCCTGACCTTGGGCAGTCTAGTGTTC
CTGCATTCTAGTCCCTGCTGTGCTGCAGGACTTTGGGCAAGTGACCTGCCCTCTGTGAGC
CTCCCTCTGACACAGAGGAGGTGGCTCCCTTCCCCACACCTTAGAGTGGCTGGGAGGGT
AACAAAGAGGGCTGCCCCCTTTAGTCTCCTGCACCCCTGCCCCCTGGTTTACCAGAGGGA
GCGGATGAAGGATGGCAGCATCTACATGCCCCATCACCAACTCTGAGGCACCTGGGGT
[G, A]
GGGGGCGGAGCCCAGGCCTCTGGCTGCTCCCTGTGGGAGCCATTGGAATGTATCCCTG

FIGURE 3F

ACAGGCCCCCTTCGCTCCACCTCAACCCAGGTCTTGGATTTTCAGGTCCCTCCACCCC
ATTCTGAGTCTCTGTCCTTCTCCTTCCACCCGCTCCCAGGGTTTCCCACCACAGGGTCTG
GAAGTGTGTGTGACGCCCATTGAGCTGTTACCCGAAGTCAGATTAAAAATCAGGGAGTGT
TTTCCCTCGTTTCTGTACCAAGGTGTTGGCTCCATTCCCTCATGGTAGGAGGGGAGGGGTC

12087 GAGCCTGTGTGTTGAGGGCAGGACTGGTCTCTGTCCGTGGTGCTGACCCACCAGCCACTT
CCAGGAAAGATGGGGCTGCCTGGCAAGGTTGGCTGAGCCTCAAAAGAGGAAGCCTCTCTC
ACCACCAACTCCTTCCCTTCTAGTCCCCATCTCCTCCAGTGGGATAACATCTGAAGCTATA
CCTCCCCGCACCACCACAGTCC TGGAGTGAGGGACTCAAGAAGCTGGGGGGCAGGGGGAG
GCAGGTTCAGTGGTTCACATCTTTAATCCCACTGCTTTGGGAGGCCAAGGCAGGAGGATC
[A,G]
CTTGAGGCCAGCCTGGACAACATAGTAAGAC

FIGURE 3G